p. 119-120+122+124+126+128+131 = (7)
p. d. 11-1996

Improving furnaces with model-based temperature control

Cole Porter, Allan Laser, *Silicon Valley Group Inc., Thermco Systems Division, San Jose, California*
Jim Hertle, *Motorola COM-1, Phoenix, Arizona*
Pradeep Pandey, Mark Erickson, Sunil Shah, *RelMan Inc., Santa Clara, California*

Thermocouples (t/c) positioned in the furnace element and process chamber of an oxidation/diffusion or LPCVD furnace system are used as inputs for adjusting the power d livered to multizone heating elements. Until recently, the industry relied on proportional integral derivative (PID) algorithms to indirectly control the temperature of the wafer based upon input from the thermocouples.

Model-based temperature control (MBTC) uses a mathematical model of the thermal characteristics of the furnace, process chamber, and wafer stack to predict and control the wafer's temperature. This article describes the implementation, testing, and results of MBTC on a vertical thermal reactor system. Significant benefits result from improved uniformity, run-to-run repeatability, reliability, and cycle-time reduction for selected oxidation/diffusion processes.

In the fabrication of semiconductor devices, oxidation and diffusion furnaces elevate the silicon substrates to a sufficient temperature to carry out diffusion of a dopant or oxidation on the surface. Typically, oxidation and diffusion furnaces are configured in a similar manner with the same reactant gas and temperature capabilities, allowing most of these systems to perform the same tasks in a production line. Process temperatures range from 400–1250°C.

For a desired oxidation or diffusion process, wafers are loaded into a quartz fixture called a "boat," as shown in Fig. 1. The boat is then inserted or "pushed" slowly into the furnace tube which is controlled to a temperature above room temperature but low enough to prevent thermal stress damage to the wafers during the insertion. Push temperatures in this paper range from 650–800°C. After boat insertion, the furnace temperature is ramped to the desired diffusion or oxidation temperature, and allowed to stabilize; then the oxidation or diffusion step is executed in an appropriate gas ambient. An inert gas purge or anneal sequence then occurs, typically at the same temperature, followed by a temperature ramp down to the desired wafer extraction or "pull" temperature. The pull is normally performed at the same temperature as the push, and the wafers are slowly removed from the furnace tube into the room temperature environment.

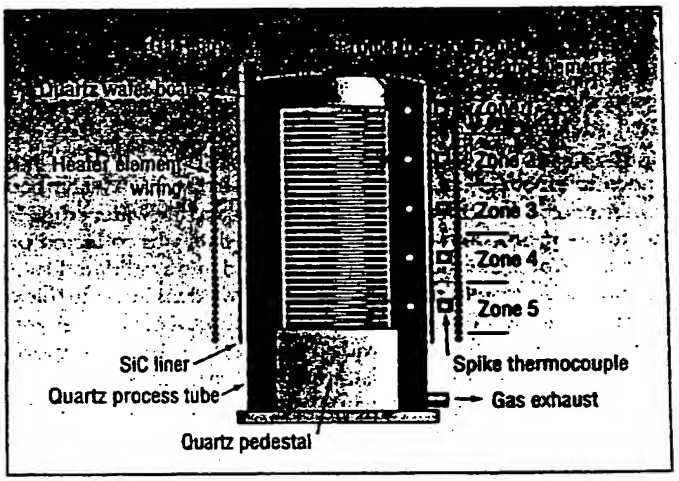

Figure 1. Schematic of a typical vertical thermal reactor (VTR) showing the location of spike and profile thermocouples. Neither set directly measures the wafers, and the discrepancy between measured and wafer temperatures results in reduced process control.

The general reactor configuration includes a process chamber made up of a high-purity quartz process tube, wafer boat, and pedestal, which the wafer boat resides on (Fig. 1). A hydrogen combustion system is housed in an externally configured heating element. The process chamber is positioned vertically inside a helically wound heater element with five control zones. A SiC liner is positioned between the process chamber and the heater element. Gas injection is accomplished via mass flow controllers (MFCs) through a manifold assembly that directs process gases into the top of the process tube.

As device geometries continue to shrink, chip manufacturers must achieve proportional improvements in the control of gate oxide uniformities, line widths, diffused junction depths, and

related dimensions. Comparable technology advances must be realized in diffusion furnace temperature control.

Furnace temperature is monitored and controlled via thermocouples located strategically to represent each of the five control zones of the heater element. "Spike" thermocouples are positioned outside of the process chamber in close proximity to the element windings in each of the five c ntrol zones. Readings from these spike thermocouples approximate the heater element temperature. "Profile" thermocouples are positioned inside the process chamber in close proximity to the wafer stack, one in each of the five control zones. These profile thermocouples approximate the wafer stack temperature.

A vertical diffusion furnace may process as many as 150 production wafers in a single batch. To meet the needs of next generation device geometries, the entire batch must have tighter thermal uniformity and lower expenditure of thermal budget during ramps and stabilization. MBTC technology successfully addresses these issues and provides the wafer temperature control required to meet present and future batch manufacturing needs of the semiconductor industry. This article describes the results of a joint development agreement (JDA) between Silicon Valley Group's (SVG) Thermco Division and Motorola's COM-1 to implement MBTC on a vertical furnace.

Evaluation methods

Two metrics were used to evaluate the performance of temperature control systems:

1) direct measurement of wafer temperature or "thermal uniformity"and
2) oxide thickness uniformity.

Thermal uniformity results were obtained using type R thermocouple instrumented wafers from SensArray. Each wafer had a thermocouple located at the wafer center and the wafer edge. In addition, profile and spike temperature data was collected on all runs. Wafer thermocouple data was fed into a validated two-stream oxide growth model to compute variations in oxide thickness for each wafer thermocouple location. This model was used to compute the target oxide thickness, D_a, using the setpoint temperature, and actual oxide thickness, D_i, using measured wafer temperatures. Using these quantities, two measures of oxide uniformity were computed:

$$\sigma_{ox-setpoint} = \sqrt{\frac{\sum_{i=1}^{n}(D_a - D_i)^2}{n-1}}\,(\text{Å}) \qquad (1a)$$

and

$$\sigma_{ox-self} = \sqrt{\frac{\sum_{i=1}^{n}(D_s - D_i)^2}{n-1}}\,(\text{Å}) \qquad (1b)$$

where

n = the number of measurement locations considered

D_s = the mean of all D_i.

Computed oxide variations were converted to equivalent temperature variations using sensitivity of oxide thickness to temperature, S_{ox} (Å/°C), computed from the oxide growth model. Thermal uniformity is computed two ways. Self-uniformity is computed with respect to its own mean. Total thermal uniformity (TTU) is computed for all wafer thermocouple locations in a batch with respect to the value for the temperature set-point trajectory. Thus,

$$\text{TTU} = \frac{\sigma_{ox-setpoint}}{S_{ox}}\,(°\text{C}) \qquad (2a)$$

and

$$\text{Self thermal uniformity} = \frac{\sigma_{ox-self}}{S_{ox}}\,(°\text{C}) \qquad (2b)$$

TTU values are larger than self-uniformity numbers. Smaller TTU values ensure that run-to-run variations and furnace-to-furnace variations are small.

Oxide thickness uniformity tests were performed using lightly boron doped <100> crystal orientation test wafers. All processing was performed using 150-mm wafers and evaluated on an ellipsometer using a nine-point measurement pattern with 3-mm edge exclusion. Oxide uniformity results are reported as % 1-σ values calculated by dividing 1-σ Å values by the average thickness x_a.

$$1\text{-}\sigma\ \text{Å} = \sqrt{\frac{\sum_{i=1}^{n}(x_a - x_i)^2}{n-1}}\,\text{Å} \qquad (3)$$

The wafer load for all runs was based on the Motorola COM-1 production load size of 37 wafers with a wafer spacing of 0.517 in. The thermal uniformity runs had 5 t/c wafers placed in the load at wafer positions 6, 12, 19, 26, and 32 (counting from the top down); the rest of the load was made up of filler wafers. The t/c wafer runs were performed in a nitrogen ambient to prevent oxidation of the SensArray wafers. The t_{ox} uniformity runs had test wafers loaded in positions 6, 19, and 32.

Process descriptions

Three Motorola oxidation and diffusion processes were used to evaluate and compare PID vs. MBTC. These processes were chosen due to their inherent sensitivity to temperature control, which greatly influences oxide thickness and diffused junction depth uniformity. The process recipes were transferred to the lab furnace and implemented in both PID and MBTC temperature control versions. The 200 Å and 350 Å processes were standard oxidati n cycles which ramp up to oxidation temperature and stabilize in an inert gas, oxidize in either dry O_2 or steam, perform a post-oxidation anneal, and ramp down to the pull temperature in an inert gas. All processing was performed at atmospheric pressure. A production emitter anneal process is normally performed entirely in an inert gas to diffuse the emitter junction to a desired and repeatable depth within the base region. This process was modified for this project by replacing the inert gas with O_2 throughout the entire process, including the push and pull. In this way, the time-averaged thermal uniformities could be evaluated by measuring oxide thicknesses.

PID control limitations

The industry standard PID control algorithm adds three terms together to determine power applied to each zone of the heater element. Each term is a function of only the error between the measurement and the setpoint for the control loop. By design, a PID controller can only act on temperature measurements local

continued on page 122

Furnace MBTC continued from page 120

Table 1. Comparison of PID and MBTC performance results

Process	t_{ox} Uniformity (% 1-σ)	Throughput (wafers/hr)
200 Å, 1000°C dry ox		
PID	0.59	6.5
MBTC	0.59	7.7
350 Å, 900°C wet ox		
PID	1.3	11.3
MBTC	0.7	13.2
Emitter anneal, 950°C		
PID	1.67	11.6
MBTC	0.77	11.6

to the zone being controlled by it.

$$\text{Error } e = \text{Setpoint} - \text{Measurement}$$

$$\text{Heater power} = K * e + K_d * de/dt + K_I * \int e\,dt \qquad (4)$$

The (first) proportional term acts as a primary negative feedback on the setpoint error. The (second) derivative term acts as a brake to reduce speed of response, and the (third) integral term acts to reduce steady-state offset in the error.

PID controllers have been popular over the last 40 years because they could be implemented using analog components. When microprocessors came along, digital PIDs were simple enough to be coded and implemented with the modest computation resources available on 8-bit processors. PIDs are tuned using rules of thumb, such as the Zielgler-Nichols rules, directly from observed thermal step responses on the furnace.

Spike, Profile, Ratio, and Cascade control strategies are used on production furnaces. Spike and Profile strategies use the spike and profile t/c measurements in forming the setpoint error. Ratio control uses spike and profile t/c signals in a fixed ratio to form the measurement signal. Cascade control refers to a series or cascade of two PID loops per zone. In addition, programmable limits are placed on power outputs.

Tuning rules for PID assume there is no coupling among heating zones. However, the profile thermocouple in the end-zone is affected by power applied to the adjacent zone. This coupling makes tuning difficult and places inherent limitations on performance. For example, tightening cascade control or profile control to reduce offsets leads to oscillations. A spike thermocouple sees less effect of power in adjacent zones than the profile thermocouple because the spike thermocouple is closer to the heater element. Therefore, spike PID control can be tuned to respond much faster than profile or cascade control. Spike control is used during stabilization and processing to ensure fast PID response.

After profiling, the steady-state offset between profile and spike readings is estimated and stored in a profile table. This offset is used to bring steady-state profile temperatures to the processing temperature by adding a feed-forward bias to the spike control loop.

In practice, furnace reprofiling takes up to 12 hours and can be required as often as once a week. In some processes, where cycle time and thermal budget permit, a real-time profiling step of up to 30 min is added to the recipe at the process temperature.

PID performance baseline

Process performance relative to the current Motorola performance results was verified using standard PID temperature control. Each PID process was then baselined by performing passive data collections (PDCs) consisting of five consecutive runs each. Passive data collection is performed with a strict "hands off" approach where no adjustments are made to the process or system throughout all runs. The t_{ox} uniformity and throughput performance results for the PID baseline are depicted in Table 1. The % 1-σ is calculated using all points from the five runs, with each PDC having 135 data points.

Figure 2 plots temperature deltas for the SensArray t/c wafers vs. the profile thermocouples. This plot displays temperature delta vs. time from the start of ramp up through the oxidation or anneal step to the end of the ramp-down step. Two main observations are evident:

1) profile t/c temperature is not an accurate approximation of wafer temperatures, and
2) PID temperature control does not take into account wafer temperature. Even with finely tuned PID, wafer to profile t/c temperature deltas will always be present, since true wafer temperatures are unknown to PID control.

The largest deltas, up to 15°C, exist during the ramp up (from 42 to 62 min) and ramp down (from 95 min to the end of the chart). During the oxidation or anneal (steady-state) step the wafer temperatures are within 5°C of the respective profile thermocouples (Fig. 3). For this process, PID shows good control

continued on page 124

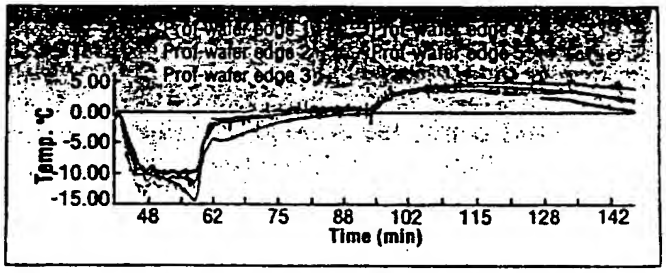

Figure 2. The difference between profile thermocouple and wafer edge thermocouple for each zone shows large differences during the ramp-up and ramp-down (dynamic) measurements.

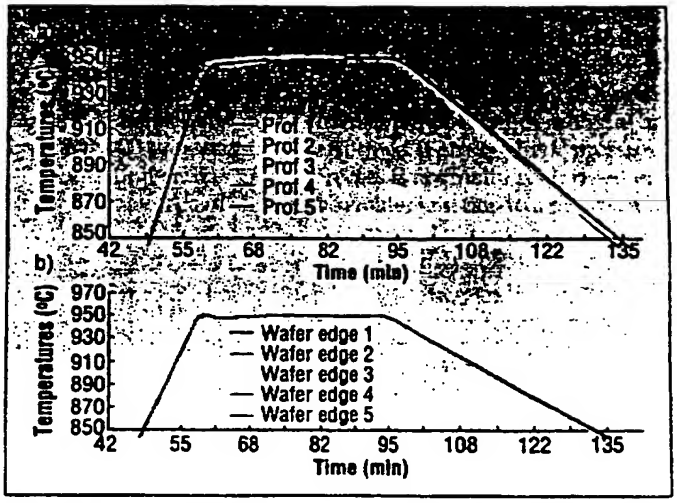

Figure 3. The a) profile and b) wafer-edge temperatures show good control under PID for the anneal process. The profile temperature undershoot at the end of the ramp up keeps the wafer temperature from overshooting.

Furnace MBTC continued from page 122

of the wafer temperatures to the temperature setpoint. The undershoot at the end of the ramp keeps the wafer temperature from overshooting.

MBTC furnace configuration

MBTC was implemented on a vertical thermal reactor (VTR) model VTR7000 located in the Thermco Division's applications laboratory in San Jose, CA, as well as on a production VTR7000 located in Motorola's COM-1 facility in Phoenix, AZ. All development work and performance evaluation was performed on the SVG applications lab VTR. The process chamber and element configuration of the VTR in the lab is a duplicate of the VTR configuration at COM-1, facilitating a smooth MBTC technology transfer to the COM-1 VTR.

The term Model-Based Control is used to denote a broad class of control approaches using systematic process modeling to determine control algorithms. In the past, practical applications of Model Based Control have been limited to control of equipment variables measured on-line, such as the profile and spike thermocouples.

The MBTC approach described here is unique in the use of models to create production-worthy *virtual sensors* for wafer temperatures and silicon parameters and in the use of these sensors for real-time control. RelMan's Thermal Reactor Optimizer-I (TRO-I) software uses an advanced mathematical thermal model of wafers and the wafer environment (furnace/process chamber) to estimate and control within wafer and wafer-to-wafer temperatures. It uses existing spike and profile thermocouples as well as heater powers to control heaters in real-time.

In order to implement MBTC on the VTR, several hardware and software changes were required. These included upgrading the CPU used for temperature control, incorporating software to perform MBTC, adding hardware to measure heating element power in each zone, and upgrading the user-interface software. One of the requirements of the MBTC algorithm is measurement of the actual power being delivered to the element. The standard VTR configuration does not have furnace elem nt power measurement capabilities. Additional hardware was thus designed into the VTR to sense current and voltage for each zone and report a power value back to the MBTC CPU board. MBTC parameters were added to the data collection capabilities and furnace configuration menus.

Oxidation control

In most oxidation processes, wafer temperature adherence to the process setpoint is most important during the oxidation step. In particular, for optimum oxide thickness uniformities and oxide integrity, it is more critical to control wafer temperature during and immediately after the oxidation than during the ramp up and ramp down. In this type of process we incorporate a MBTC strategy called flat trajectory, in which the model drives the wafer temperatures to the process setpoint and holds them at setpoint for a specified time. The objective is to tightly control wafer temperatures during oxidation and anneal for the full batch of processed wafers. In submicron oxidation processes, reducing ramp and stabilization time significantly reduces expenditure of thermal budgets.

Diffusion control

For a diffusion process, wafer temperatures must be precisely controlled through the entire thermal cycle. Total thermal uniformity

continued on page 126

Furnace MBTC continued from page 124

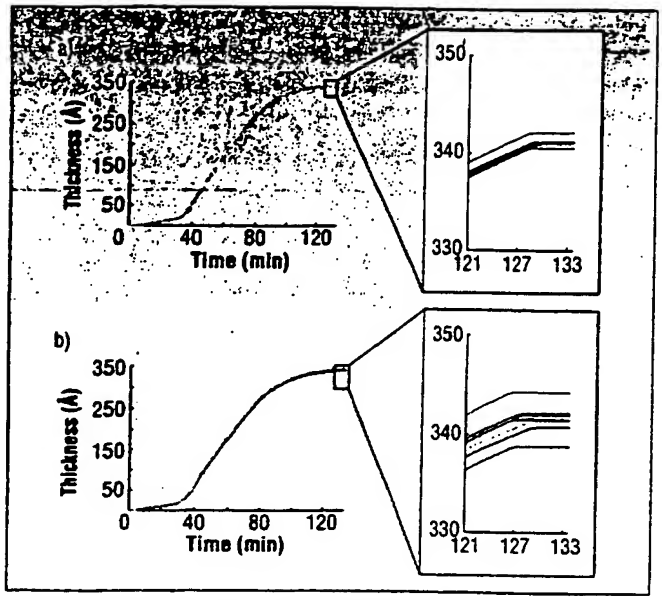

Figure 4. Calculated oxide growth as a function of time for each wafer thermocouple location during the anneal process: a) MBTC oxide growth trajectory, and b) PID oxide growth trajectory. The MBTC dynamic optimization strategy minimizes variation at the end of the process, resulting in an order of magnitude improvement in TTU over baseline PID.

(TTU) of the entire batch during ramp up, stabilization, and ramp down is critical to the control of junction depths and channel widths. Using a dynamic optimization MBTC strategy, the model drives the wafers to the appropriate temperatures to reach a low TTU value such that, at the end of the diffusion cycle, all wafers have been subjected to the same thermal budget. The model adjusts and optimizes for temperature nonuniformities across each wafer and from wafer-to-wafer in the boat. For example, the bottom-most wafers in the stack may be nominally cooler than the top-most wafers during the ramp up. The model achieves good

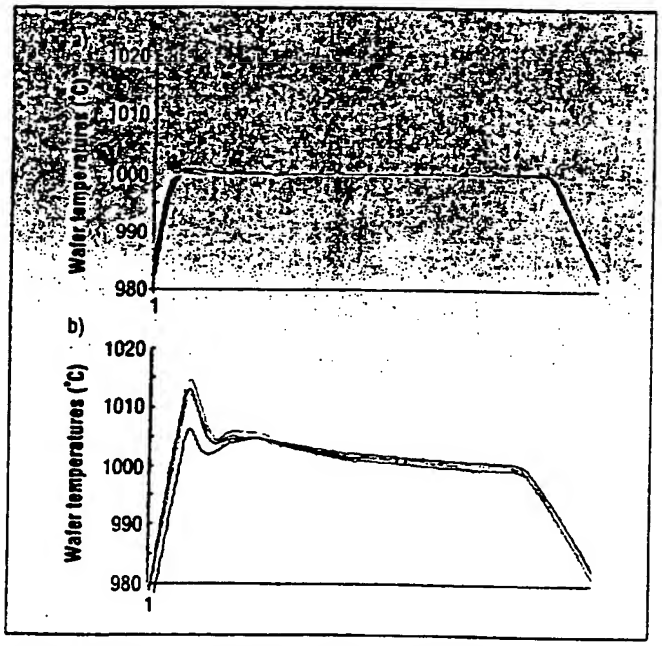

Figure 5. Wafer temperatures for a) model-based 200-Å oxidation, and b) PID 200-Å oxidation.

TTU by driving the bottom wafers sufficiently hotter than the top wafers at the end of the ramp cycle and into the drive-in step.

MBTC process optimization and results

Both modeling approaches use furnace fingerprinting to provide the empirical inputs used in achieving optimal TTU. The fingerprinting operation serves two important objectives. First, it characterizes the unique thermal behavior of the furnace configuration. Second, it verifies the integrity of all thermal subsystems of the furnace. For fingerprinting, wafer mounted thermocouples, such as SensArray wafers, are processed through the furnace in an inert atmosphere. To provide a precise and accurate model, the finger printing data must include heater powers, wafer temperatures, profile temperatures, spike temperatures, and gas flows.

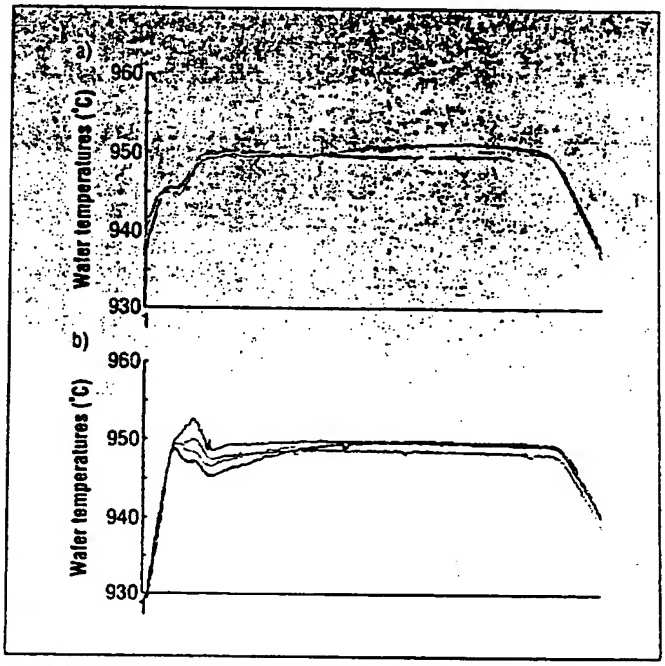

Figure 6. Wafer temperatures for a) model-based DO emitter anneal process, and b) PID emitter anneal process.

A specially designed fingerprinting kit is used during final test at the factory which includes the quartz boat and SensArray wafers. Using the fingerprinting data, a customized mathematical algorithm is created for each furnace. This allows process engineers to establish recipes that are relatively independent of different tools. Because of unavoidable differences among furnaces, a production PID controller requires furnace-specific adjustments to recipe setpoints. By eliminating these adjustments, MBTC makes recipe transfer easier. MBTC also eliminates the profiling that is required with PID control.

The MBTC version of each process was optimized to achieve the desired goals. The flat trajectory model was used for the two oxidation processes because the goal was to maximize throughput. Temperature ramp rates could be increased and stabilizations reduced or eliminated. For the 200 Å dry oxidation, the ramp-up and stabilization time was reduced from 68 to 50 min. The ramp-down time was also reduced from 117 to 102 min for a total cycle-time reduction of 33 min. For the 350-Å wet oxidation, the ramp- up and stabilization time was reduced from 44 to 25 min.

continued on page 128

Furnace MBTC continued from page 126

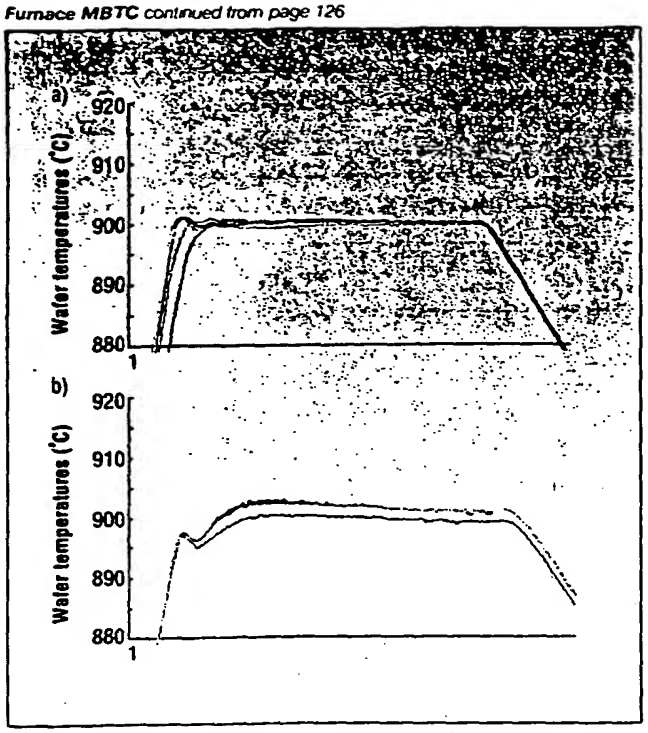

Figure 7. Wafer temperatures for a) model-based 350-Å oxidation, and b) PID 350-Å oxidation.

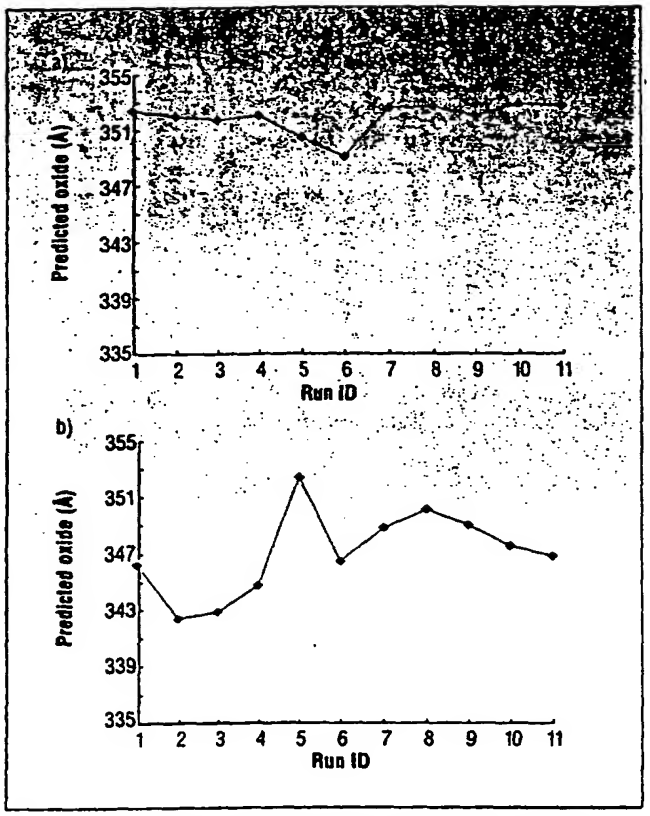

Figure 8. Comparison of predicted oxide mean thicknesses calculated from wafer temperature measurements made during the Phase I Iron Man testing: a) TRO-I anneal run means, and b) PID anneal run means. See Table 2 for corresponding run ID numbers.

The emitter anneal cycle time was not modified because the goal was a 50% improvement in total uniformity. To improve thermal uniformity, the dynamic optimization strategy was invoked during the critical anneal step. The oxide growth during this step was computed using measured wafer temperatures in a two-stream validated oxide growth model (Fig. 4). The improvement over baseline PID is 7× with Self-Mean calculations. The Total Thermal Uniformity (TTU) improvement was 18×.

Figures 5, 6, and 7 compare measured wafer temperatures between MBTC and PID control for each process. In every case MBTC yielded much tighter control of wafer temperatures relative to the desired setpoint. The deviation from setpoint in the MBTC emitter anneal process is an expected result of the Dynamic Optimization strategy. Once the MBTC version of each process was optimized, PDCs were performed to evaluate performance (see Table 1). All of the JDA goals were met or exceeded in implementing MBTC.

Measured oxide thickness uniformity and throughput were equal or better with MBTC. In particular, the throughput improvements for the two oxidation processes correspond to approximately 15% lower cost of ownership (calculated as cost/good wafer/year using the SEMATECH COO model).

MBTC reliability testing (Iron Man)

Once the capabilities of MBTC were proven, an Iron Man was performed to challenge the reliability and robustness of the system. The Iron Man is Motorola's standard procedure for testing unproven or recently developed semiconductor capital equipment. In Phase I of the Iron Man testing, various process and equipment changes (Table 2) were made to the system to strain the MBTC hardware and software and determine failure modes. In Phase II, the VTR was run for an extended period of time without any changes to test the reliability of the entire system and simulate the manufacturing environment.

Table 2. Equipment and process changes made during the Pre-Iron Man testing

Run ID #	Equipment changes
1	Profile t/c removal and replacement
2	Quartz process tube change
3	Spike t/c removal and replacement
4	Quartz pedestal change
5	Power acquisition module inversion
Run ID #	**Process changes**
6	Thermocouple amplifier PCB change
7	Gas flow change (+100% of nominal)
8	Gas flow change (50% of nominal)
9	Increased wafer load by 24
10	Increased wafer load by 48
11	Increased wafer load by 74

After each change was introduced, the furnace was tested with both MBTC and PID processes using the t/c instrumented wafer load. Wafer temperatures were analyzed to evaluate the effect of the equipment or process change on each temperature control method. Throughout the Phase I Iron Man testing, MBTC

continued on page 131

Furnace MBTC continued from page 128

performed reliably and proved to be less sensitive to equipment changes than PID. PID control degraded sufficiently after the quartz pedestal change to cause an abort for profile t/c temperature limits "out of tolerance" which was set at ±5.0°C. The system had to be reprofiled after this change to bring PID control back into specification. In addition to the abort, PID control exhibited significant run-to-run variation compared to MBTC. Figure 8 compares predicted oxide means calculated from the wafer temperature measurements made during the Phase I Iron Man testing. The predicted oxide means include measurements from all 5 t/c wafers.

During the Phase II Iron Man, the primary objective was to evaluate the reliability of the new MBTC hardware and software. The final system configuration was tested for 20 days under typical production line conditions. The laboratory furnace was run

MBTC can control actual wafer temperatures instead of relying on indirect measurements from thermocouples positioned in the furnace element and process chamber.

24 hours/day, six days/week, utilizing all three processes and both MBTC and PID control. During idle periods, the MBTC hardware and software was still operational, maintaining furnace standby temperatures. Over the course of the Phase II Iron Man, 122 process runs were completed. The MBTC hardware and software experienced no failures and did not cause any system downtime.

Conclusion

MBTC provides precision wafer temperature control in an oxidation/diffusion or LPCVD furnace. By using a thermal model of the wafers and wafer environment, MBTC can control *actual* wafer temperatures instead of relying on indirect measurements from thermocouples positioned in the furnace element and process chamber. Beyond simply controlling the furnace element temperature, MBTC can precisely predict and control the product temperature.

We have successfully implemented MBTC on a VTR7000. MBTC has improved silicon wafer thermal uniformity as much as 7× and improved oxide film thickness uniformities by more than 50%, while achieving a 15% increase in throughput and corresponding reduction in cost of ownership. The reliability and production worthiness of the new system was demonstrated by the successful completion of an Iron Man test. ■

Acknowledgments

The authors would like to acknowledge the following: Thermco Division's mechanical and electrical engineering departments and the supporting infrastructure of the organization; Motorola COM-1, SVG's first customer and end-user of MBTC on the VTR7000 furnace; Jeff Kowalski, president of SVG, Thermco Systems Division, John Gragg of Motorola's Manufacturing Technology Group, and Henry Becker of Motorola's Microprocessor and Memory Technologies Group for their vision of the need of MBTC in the industry and their leadership in the JDA; Chris Palmer for his major contribution and participation as a Motorola project manager and process engineer; and the Thermco Division Applications Laboratory Team for their contributions and support throughout the project.

COLE PORTER is a senior engineering program manager and MBTC product manager at the Thermco Systems Division of the Silicon Valley Group, where he is responsible for the implementation of MBTC technology. For the past seven years, he has held various process engineering positions at Thermco in furnace products. He has 16 years of experience in the semiconductor industry. *Silicon Valley Group, 2240 Ringwood Ave., San Jose, CA 95131; ph 408/434-0500, fax 408/325-8898.*

ALLAN LASER received his BS and MS degrees in electrical engineering from the University of British Columbia in Vancouver, Canada. He is a principal process engineer in the Process Technology Department at Silicon Valley Group, Thermco Systems Division. He has five years of experience in semiconductor manufacturing.

JIM HERTLE received his BS degree from Central Michigan University and is the equipment engineering manager for Diffusion, Implant and Thin-Film areas at Motorola's COM-1 wafer fab. Before joining Motorola, he was a mechanical engineer at Advanced Semiconductor Materials and General Signal Thinfilm Co.

PRADEEP PANDEY received his doctorate in electrical engineering from the University of California, Santa Barbara, and is the VP of R&D at RelMan Inc. He has more than 10 years of experience in advanced control and CAD tools product development.

MARK ERICKSON received his MS degree from Stanford University in 1989 and his BS and PhD degrees in 1988 and 1994 from the University of California, Santa Barbara, all in electrical engineering. He is a research scientist at RelMan, where he applies advanced control and optimization techniques to semiconductor manufacturing.

SUNIL SHAH received his doctorate in electrical engineering from Stanford University. He is president and cofounder of RelMan Inc., and has more than 17 years of experience in advanced control of equipment and processes.